

MAXIMUM LIKELIHOOD DEGREE OF VARIANCE COMPONENT MODELS

ELIZABETH GROSS, MATHIAS DRTON AND SONJA PETROVIĆ

ABSTRACT. Most statistical software packages implement numerical strategies for computation of maximum likelihood estimates in random effects models. Little is known, however, about the algebraic complexity of this problem. For the one-way layout with random effects and unbalanced group sizes, we give formulas for the algebraic degree of the likelihood equations as well as the equations for restricted maximum likelihood estimation. In particular, the latter approach is shown to be algebraically less complex. The formulas are obtained by studying a univariate rational equation whose solutions correspond to the solutions of the likelihood equations. Applying techniques from computational algebra, we also show that balanced two-way layouts with or without interaction have likelihood equations of degree four. Our work suggests that algebraic methods allow one to reliably find global optima of likelihood functions of linear mixed models with a small number of variance components.

1. INTRODUCTION

Linear models with fixed and random effects are widely used for dependent observations. Such mixed models are typically fit using likelihood-based techniques, and the necessary optimization problems can be solved using the numerical methods implemented in various statistical software packages, as discussed, for instance, in [Far06]. Such software typically takes into account that the variance parameters are nonnegative. However, general-purpose optimization procedures do not give any guarantees that a global optimum is found; compare Section 1.8 in [Jia07]. It can thus be appealing to compute maximum likelihood (ML) estimates algebraically. Since linear mixed models have rational likelihood equations, this involves careful clearing of denominators and applying symbolic and specialized numerical techniques to determine all solutions of the resulting polynomial system. An explanation of what we mean by careful clearing of denominators is given in [DSS09, Chap. 2]. While solving likelihood equations algebraically may not be feasible in large models with several random factors, modern computational algebra does allow one to fully understand the likelihood surface in practically relevant settings.

The main contribution of this paper is a study of the algebraic complexity of ML estimation in the unbalanced one-way layout with random effects. This model concerns a collection of grouped observations

$$(1) \quad Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, \dots, q, \quad j = 1, \dots, n_i.$$

The overall mean $\mu \in \mathbb{R}$ is a fixed (‘non-random’) but unknown parameter. The random effects α_i and the error terms ε_{ij} are mutually independent normal random

Key words and phrases. Analysis of variance, linear mixed model, maximum likelihood, restricted maximum likelihood, variance component.

variables. More precisely, $\alpha_i \sim \mathcal{N}(0, \tau)$ and $\varepsilon_{ij} \sim \mathcal{N}(0, \omega)$, where τ and ω denote the common variances of the random effects and the error terms, respectively. Clearly, the distribution of observation Y_{ij} is $\mathcal{N}(\mu, \tau + \omega)$, and two observations Y_{ij} and Y_{ik} from the i th group are dependent with covariance τ . A detailed discussion and examples of applications of this specific model can be found, for instance, in Chapter 3 of [SCM92] and in Chapter 11 of [SO05].

The covariance matrix of the joint multivariate normal distribution for all Y_{ij} defined by (1) is the product of the scalar ω and a matrix that is a function of the variance ratio $\theta = \tau/\omega$. Therefore, when θ is known, the likelihood equations for μ and ω are of the type encountered in generalized least squares calculations, with a unique solution that is a rational function of the data and the known value of θ . We may thus eliminate μ and ω from the likelihood equations, which then reduce to a single univariate equation. Before turning to a first example, we remark that we always tacitly assume suitable sample size conditions to be satisfied such that ML estimates exist. In particular, we assume there to be $q \geq 2$ groups with at least one group of size $n_i \geq 2$. A definitive answer to the existence problem in linear mixed models is given in [DM99] who also treat restricted maximum likelihood (REML) estimation; see [MN89] for an introduction to this technique.

Example 1. Textbook data from [DG72, §6.4] give the yield of dyestuff from 5 different preparations from each of $q = 6$ different batches of an intermediate product; the data are also available in the R package `lme4`. The layout is balanced, that is, all batch sizes are equal, here $n_i = 5$. In this case, the likelihood equations are well-known to be equivalent to a linear equation system, and the ML estimators are rational functions of the observations Y_{ij} . In terminology we will use later on, balanced one-way layouts have ML degree one. Exactly the same is true for REML.

A different picture emerges in the unbalanced case, when the batch sizes are not all equal. For illustration, we remove the first, second and sixth observation from the data. The first batch then only comprises $n_1 = 3$ preparations, and the second batch only $n_2 = 4$. The remaining batches are unchanged with $n_i = 5$ for $i \geq 3$. In this unbalanced case, the solutions of the likelihood equations correspond to the solutions of the polynomial equation

$$(2) \quad -245488320000 \theta^7 - 277109078400 \theta^6 - 58814614680 \theta^5 + 54052612853 \theta^4 \\ + 37792395524 \theta^3 + 10086075110 \theta^2 + 1279832076 \theta + 64175517 = 0.$$

As the large integers suggest, this equation is exact given the input. However, the measurements that enter the computation are, of course, rounded.

Numerical optimization using the R package `lme4` yields a local maximum of the likelihood function that corresponds to $\theta \approx 0.5585$. We may check whether this local maximum is unique, or at least a global optimum, by finding all roots of the above univariate polynomial. This is a task that can be done reliably in computer algebra systems. However, such a computation is not needed here. The polynomial in (2) has exactly one sign change in its coefficient sequence. Hence, by Descartes' rule of signs, it has precisely one positive real root. The mere construction of the polynomial thus reveals that the local maximum we computed is the unique local (and global) maximum of the likelihood function; recall that the parameter θ is restricted to be nonnegative.

A similar story unfolds for REML estimation. The only difference is that the degree of the relevant polynomial drops to five:

$$(3) \quad -17047800000\theta^5 - 6811774200\theta^4 + 5505084700\theta^3 \\ + 4048254212\theta^2 + 897954164\theta + 67458244 = 0.$$

We note that equations (2) and (3) cannot be solved by radicals. The Galois groups are the symmetric groups S_7 and S_5 , respectively.

It is natural at this point to ask for the *maximum likelihood degree* of the one-way layout as a function of the number of groups q and the group sizes n_1, \dots, n_q . The ML degree is the number of complex solutions to the likelihood equations when the data are *generic*. Indeed, the number of complex solutions is constant with probability one, and a data set is generic if it is not part of the null set for which the number of complex solutions is different. The REML degree is defined in just the same way, but starting from different equations. Either degree measures the algebraic complexity of the computation of the estimates. In Example 1, it is simply the degree of a univariate polynomial in $\theta = \tau/\omega$ whose roots yield the possibly complex vectors (μ, ω, τ) that solve the likelihood equations. For more background on ML degrees, see [HKS05, CHKS06, BHR07, DSS09, Stu09, HS10].

Our main result answers the above question. Theorem 1, which we prove in later sections, gives formulas for both the ML and the REML degree of possibly unbalanced one-way layouts and offers a direct comparison of the algebraic complexity of the two approaches. The theorem is conveniently stated using a notion of multiplicities. Suppose $v = (v_1, \dots, v_q) \in \mathbb{Z}^q$ is a tuple of integers. If v has M distinct entries, then the multiplicities of v form the integer multiset $\{m_1, \dots, m_M\}$, where m_j counts how often the j th distinct entry of v appears among all entries of v .

Theorem 1. *Consider a one-way layout with random effects for q groups that are of sizes n_1, \dots, n_q . Suppose M of the group sizes are distinct, with associated multiplicities m_1, \dots, m_M . Let $M_2 = \#\{j : m_j \geq 2\}$. Then the ML degree is $3M + M_2 - 3$, and the REML degree is $2M + 2M_2 - 3$. The ML degree exceeds the REML degree unless $M_2 = M$, in which case equality holds.*

The condition $M_2 = M$ holds if each group size appears at least twice. In the balanced case, we have $M = M_2 = 1$ and the theorem recovers the well-known fact that both degrees are one; compare [Hoc85, SCM92, SO04]. Each degree is maximal when the group sizes n_1, \dots, n_q are pairwise distinct. The degrees are then $3q - 3$ for ML and $2q - 3$ for REML.

Example 2. The model for the dyestuff data from Example 1 has $q = 6$ groups. The unbalanced case we considered had group sizes $(n_1, \dots, n_6) = (3, 4, 5, 5, 5, 5)$. The multiplicities are $\{1, 1, 4\}$. Our formulas confirm the ML and REML degree to be $3 \cdot 3 + 1 - 3 = 7$ and $2 \cdot 3 + 2 \cdot 1 - 3 = 5$, respectively. As another example, if $(n_1, \dots, n_6) = (4, 4, 3, 2, 2, 2)$, then the ML degree is 8 and the REML degree is 7.

The remainder of the paper is structured as follows. In Section 2, we review the derivation of the likelihood equations for ML and REML estimation. Section 3 contains the proof of the ML degree formula from Theorem 1, and Section 4 treats the REML degree. Each proof consists of a detailed study of a univariate rational equation in the variance ratio θ . In Section 5, we demonstrate that algebraic computations are feasible for more general linear mixed models. More precisely,

we treat a one-way layout with $q = 109$ unbalanced groups and a mean structure given by two covariates that is relevant in a recent application. In Section 6, we consider balanced two-way layouts. These are known to have REML degree equal to one, and we show that the ML degree is four, which means that ML estimates are available in closed form in the sense of Cardano's formula. Our conclusions are summarized in Section 7, where we also give two examples of unbalanced one-way random effects models with bimodal likelihood functions.

2. THE LIKELIHOOD EQUATIONS

Let n_1, \dots, n_M be unique group sizes with associated multiplicities m_1, \dots, m_M . Let $Y_{ij} = (Y_{ij1}, \dots, Y_{ijn_i})$ be the vector comprising the observations in the j th group of size n_i . Then the model for the one-way layout given by (1) can equivalently be described as stating that $Y_{11}, \dots, Y_{1m_1}, Y_{21}, \dots, Y_{Mm_M}$ are independent multivariate normal random vectors with

$$Y_{ij} \sim \mathcal{N}(\mu \mathbf{1}_{n_i}, \Sigma_{n_i}(\omega, \tau)),$$

where the covariance matrix is

$$\Sigma_{n_i}(\omega, \tau) = \omega I_{n_i} + \tau \mathbf{1}_{n_i} \mathbf{1}_{n_i}^T.$$

Here, $\mathbf{1}_n = (1, \dots, 1)^T \in \mathbb{R}^n$, and I_n is the $n \times n$ identity matrix.

2.1. Maximum likelihood. Ignoring additive constants and multiplying by two, the log-likelihood function of the one-way model is

$$(4) \quad \ell(\mu, \omega, \tau) = \sum_{i=1}^M \sum_{j=1}^{m_i} \log \det(K_{n_i}(\omega, \tau)) - (Y_{ij} - \mu \mathbf{1}_{n_i})^T K_{n_i}(\omega, \tau) (Y_{ij} - \mu \mathbf{1}_{n_i}),$$

where

$$(5) \quad K_{n_i}(\omega, \tau) = \frac{1}{\omega} I_{n_i} - \frac{\tau}{\omega(\omega + n_i \tau)} \mathbf{1}_{n_i} \mathbf{1}_{n_i}^T.$$

is the inverse of $\Sigma_{n_i}(\omega, \tau)$. The inverse has determinant

$$(6) \quad \det(K_{n_i}(\omega, \tau)) = \frac{1}{\omega^{n_i-1}(\omega + n_i \tau)}.$$

Let $N = m_1 n_1 + \dots + m_M n_M$ be the total number of observations. For each $i = 1, \dots, M$, define the *group averages*

$$\bar{Y}_{ij} = \frac{1}{n_i} \sum_{k=1}^{n_i} Y_{ijk}, \quad j = 1, \dots, m_i,$$

and the average across the groups of equal size

$$\bar{Y}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} \bar{Y}_{ij}.$$

From the averages, compute the *between-group sum of squares*

$$B_i = \sum_{j=1}^{m_i} (\bar{Y}_{ij} - \bar{Y}_i)^2.$$

Note that, for generic data, $B_j = 0$ if and only if $m_j = 1$. Therefore, it suffices to consider the sums of squares B_i with $m_i \geq 2$. Finally, define the *within-group sum of squares*

$$W = \sum_{i=1}^M \sum_{j=1}^{m_i} \sum_{k=1}^{n_i} (Y_{ijk} - \bar{Y}_{ij})^2,$$

which is positive for generic data.

Proposition 1. *Upon the substitution $\kappa = 1/\omega$ and $\theta = \tau/\omega$, the log-likelihood function for the one-way layout can be written as*

$$(7) \quad \ell(\mu, \kappa, \theta) = N \log(\kappa) - \kappa W - \left[\sum_{i=1}^M m_i \log(1 + n_i \theta) \right] - \kappa \left[\sum_{i=1}^M \frac{n_i}{1 + n_i \theta} B_i \right] - \kappa \left[\sum_{i=1}^M \frac{m_i n_i}{1 + n_i \theta} (\bar{Y}_i - \mu)^2 \right].$$

Proof. Applying (5), the quadratic form in (4) can be expanded into

$$\begin{aligned} & (Y_{ij} - \mu \mathbf{1}_{n_i})^T K_{n_i}(\omega, \tau) (Y_{ij} - \mu \mathbf{1}_{n_i}) \\ &= \frac{1}{\omega} \sum_{k=1}^{n_i} (Y_{ijk} - \mu)^2 - \frac{\tau}{\omega(\omega + n_i \tau)} [(Y_{ij} - \mu \mathbf{1}_{n_i})^T \mathbf{1}_{n_i}]^2 \\ &= \frac{1}{\omega} \sum_{k=1}^{n_i} (Y_{ijk} - \bar{Y}_{ij})^2 + \frac{n_i}{\omega} (\bar{Y}_{ij} - \mu)^2 - \frac{\tau}{\omega(\omega + n_i \tau)} n_i^2 (\bar{Y}_{ij} - \mu)^2 \\ &= \kappa \frac{n_i}{1 + n_i \theta} (\bar{Y}_{ij} - \mu)^2 + \kappa \sum_{k=1}^{n_i} (Y_{ijk} - \bar{Y}_{ij})^2. \end{aligned}$$

Using this expression and (6), the log-likelihood function is seen to be equal to

$$(8) \quad \ell(\mu, \kappa, \theta) = N \log(\kappa) - \kappa W - \left[\sum_{i=1}^M m_i \log(1 + n_i \theta) \right] - \kappa \left[\sum_{i=1}^M \sum_{j=1}^{m_i} \frac{n_i}{1 + n_i \theta} (\bar{Y}_{ij} - \mu)^2 \right].$$

The claimed form of $\ell(\mu, \kappa, \theta)$ is now obtained by expanding the last sum as

$$(9) \quad \sum_{j=1}^{m_i} \frac{n_i}{1 + n_i \theta} (\bar{Y}_{ij} - \mu)^2$$

$$\begin{aligned} (10) \quad &= \frac{n_i}{1 + n_i \theta} \sum_{j=1}^{m_i} [(\bar{Y}_{ij} - \bar{Y}_i)^2 + (\bar{Y}_i - \mu)^2 + 2(\bar{Y}_{ij} - \bar{Y}_i)(\bar{Y}_i - \mu)] \\ &= \frac{m_i n_i}{1 + n_i \theta} (\bar{Y}_i - \mu)^2 + \frac{n_i}{1 + n_i \theta} B_i. \end{aligned}$$

□

The partial derivatives of the log-likelihood function from Proposition 1 are

$$(11) \quad \frac{\partial \ell}{\partial \mu} = 2\kappa \sum_{i=1}^M \frac{m_i n_i}{1 + n_i \theta} (\bar{Y}_i - \mu),$$

$$(12) \quad \frac{\partial \ell}{\partial \kappa} = \frac{N}{\kappa} - \left[W + \sum_{i=1}^M \frac{m_i n_i}{1 + n_i \theta} (\bar{Y}_i - \mu)^2 + \sum_{i=1}^M \frac{n_i}{(1 + n_i \theta)} B_i \right],$$

$$(13) \quad \frac{\partial \ell}{\partial \theta} = - \left[\sum_{i=1}^M \frac{m_i n_i}{1 + n_i \theta} \right] + \kappa \left[\sum_{i=1}^M \frac{m_i n_i^2}{(1 + n_i \theta)^2} (\bar{Y}_i - \mu)^2 + \sum_{i=1}^M \frac{n_i^2}{(1 + n_i \theta)^2} B_i \right].$$

Since $N \neq 0$, the equation system obtained by setting the three partials to zero has the same solution set as the equation system

$$(14) \quad \sum_{i=1}^M \frac{m_i n_i}{1 + n_i \theta} (\bar{Y}_i - \mu) = 0,$$

$$(15) \quad N - \kappa \left[W + \sum_{i=1}^M \frac{m_i n_i}{1 + n_i \theta} (\bar{Y}_i - \mu)^2 + \sum_{i=1}^M \frac{n_i}{1 + n_i \theta} B_i \right] = 0,$$

$$(16) \quad \kappa \left[\sum_{i=1}^M \frac{m_i n_i^2}{(1 + n_i \theta)^2} (\bar{Y}_i - \mu)^2 + \sum_{i=1}^M \frac{n_i^2}{(1 + n_i \theta)^2} B_i \right] - \left[\sum_{i=1}^M \frac{m_i n_i}{1 + n_i \theta} \right] = 0.$$

Now we can solve equation (14) for μ , substitute the result into equation (15) and solve for κ . Both μ and κ are then expressed in terms of θ . Substituting the expressions into (16), we obtain a univariate rational equation in θ . Our proof of the ML degree formula in Theorem 1 proceeds by cancelling terms from the numerator and denominator of this rational expression. This is the topic of Section 3.

2.2. Restricted maximum likelihood. The REML method uses a slightly different likelihood function that is obtained by considering a projection of the observed random array $(Y_{ijk}) \in \mathbb{R}^N$. The mean of this array has all entries equal to μ . In other words, it is modelled to lie in the space $\mathcal{L} \subset \mathbb{R}^N$ spanned by the array with all entries equal to one. The likelihood function used in REML is obtained by taking the observation to be the projection of (Y_{ijk}) onto the orthogonal complement of \mathcal{L} . The distribution of the projection no longer depends on μ and so the REML function only has (τ, ω) or, equivalently, (κ, θ) as arguments.

Using the formulas given, for instance, in [MN89], and simplifying the resulting expressions similar to what was done in the proof of Proposition 1, we obtain the following expression for the restricted log-likelihood function.

Proposition 2. *Upon the substitution $\kappa = 1/\omega$ and $\theta = \tau/\omega$, the restricted log-likelihood function for the one-way layout can be written as*

$$(17) \quad \bar{\ell}(\kappa, \theta) = (N - 1) \log(\kappa) - \kappa W - \left[\sum_{i=1}^M m_i \log(1 + n_i \theta) \right] \\ - \log \left(\sum_{i=1}^M \frac{m_i n_i}{1 + n_i \theta} \right) - \kappa \left[\sum_{i=1}^M \frac{n_i}{1 + n_i \theta} B_i \right] - \kappa \left[\sum_{i=1}^M \frac{m_i n_i}{1 + n_i \theta} (\bar{Y}_i - \hat{\mu}(\theta))^2 \right],$$

with

$$(18) \quad \hat{\mu}(\theta) = \frac{\sum_{i=1}^M \sum_{j=1}^{m_i} \frac{n_i}{1+n_i\theta} \bar{Y}_{ij}}{\sum_{i=1}^M \sum_{j=1}^{m_i} \frac{n_i}{1+n_i\theta}} = \frac{\sum_{i=1}^M \frac{m_i n_i}{1+n_i\theta} \bar{Y}_i}{\sum_{i=1}^M \frac{m_i n_i}{1+n_i\theta}}.$$

Note that $\hat{\mu}(\theta)$ is the solution to the equation in (14). Computing $\hat{\mu}(\theta)$ is the standard way to obtain an estimate of μ from a REML estimate of θ .

The partial derivatives of the restricted log-likelihood function from Proposition 2 are

$$(19) \quad \frac{\partial \bar{\ell}}{\partial \kappa} = \frac{N-1}{\kappa} - \left[W + \sum_{i=1}^M \frac{m_i n_i}{1+n_i\theta} (\bar{Y}_i - \hat{\mu}(\theta))^2 + \sum_{i=1}^M \frac{n_i}{1+n_i\theta} B_i \right],$$

$$(20) \quad \frac{\partial \bar{\ell}}{\partial \theta} = - \left[\sum_{i=1}^M \frac{m_i n_i}{1+n_i\theta} \right] + \frac{\sum_{i=1}^M \frac{m_i n_i^2}{(1+n_i\theta)^2}}{\sum_{i=1}^M \frac{m_i n_i}{(1+n_i\theta)}} + \kappa \left[\sum_{i=1}^M \frac{m_i n_i^2}{(1+n_i\theta)^2} (\bar{Y}_i - \hat{\mu}(\theta))^2 + \sum_{i=1}^M \frac{n_i^2}{(1+n_i\theta)^2} B_i \right].$$

The equation $\partial \bar{\ell} / \partial \kappa = 0$ is easily solved. Substituting the unique solution $\hat{\kappa}(\theta)$ into the equation $\partial \bar{\ell} / \partial \theta = 0$ yields again a univariate rational equation in θ . The proof of the REML degree formula in Theorem 1 requires studying cancellations from the numerator and denominator of this equation, which is the topic of Section 4.

3. PROOF OF FORMULA FOR ML DEGREE

Our proof of the ML degree formula in Theorem 1 proceeds in two steps. First, in Lemma 1 we derive a univariate rational equation whose number of zeros is the ML degree of the model. Second, we simplify it in Lemmas 2 and 3 by clearing common factors from the numerator and the denominator.

Fix the following notation, used throughout. For a vector $a = (a_1, \dots, a_M) \in \mathbb{R}^M$, define the rational functions

$$r_a(\theta) = \sum_{i=1}^M \frac{m_i n_i}{1+n_i\theta} a_i \quad \text{and} \quad s_a(\theta) = \sum_{i=1}^M \frac{m_i n_i^2}{(1+n_i\theta)^2} a_i.$$

We write $r_1, r_{B/m}, r_Y, r_{Y^2}$ for the functions r_a that have

$$a = \mathbf{1}_M, \quad a = \left(\frac{B_1}{m_1}, \dots, \frac{B_M}{m_M} \right), \quad a = (\bar{Y}_1, \dots, \bar{Y}_M), \quad a = (\bar{Y}_1^2, \dots, \bar{Y}_M^2),$$

respectively. It is clear from Section 2 that forming a common denominator for the rational equations to be studied involves the product

$$d(\theta) = \prod_{i=1}^M (1+n_i\theta) = d_1(\theta) d_2(\theta),$$

where

$$d_1(\theta) = \prod_{\{i:m_i=1\}} (1+n_i\theta), \quad d_2(\theta) = \prod_{\{i:m_i \geq 2\}} (1+n_i\theta).$$

For a vector $a \in \mathbb{R}^M$, define the degree $M - 1$ polynomial

$$f_a(\theta) = d(\theta)r_a(\theta) = \sum_{i=1}^M m_i n_i a_i \prod_{j \neq i} (1 + n_j \theta)$$

and the degree $2(M - 1)$ polynomial

$$g_a(\theta) = d(\theta)^2 s_a(\theta) = \sum_{i=1}^M m_i n_i^2 a_i \prod_{j \neq i} (1 + n_j \theta)^2.$$

Lemma 1. *The ML degree of the one-way layout is the degree of the numerator created when cancelling all common factors from numerator and denominator of the following rational function in θ :*

$$(21) \quad \frac{1}{Nd(\theta)^2 f_1(\theta)^2} \times (N [f_1(\theta)^2 g_{Y^2}(\theta) - 2f_Y(\theta)f_1(\theta)g_Y(\theta) + f_Y(\theta)^2 g_1(\theta) + f_1(\theta)^2 g_{B/m}(\theta)] - f_1(\theta)^2 [Wf_1(\theta)d(\theta) + f_{Y^2}(\theta)f_1(\theta) - f_Y(\theta)^2 + f_1(\theta)f_{B/m}(\theta)]).$$

Proof. Adopting the notation above, the solution of the first of the likelihood equations in (14) can be written as

$$(22) \quad \hat{\mu}(\theta) = \frac{r_Y(\theta)}{r_1(\theta)}.$$

Next, rewrite the following term from the system of the three critical equations:

$$(23) \quad \sum_{i=1}^M \frac{m_i n_i}{1 + n_i \theta} (\bar{Y}_i - \hat{\mu}(\theta))^2 = r_{Y^2}(\theta) - 2 \frac{r_Y(\theta)}{r_1(\theta)} \sum_{i=1}^M \frac{m_i n_i}{1 + n_i \theta} \bar{Y}_i + \frac{r_Y(\theta)^2}{r_1(\theta)^2} \sum_{i=1}^M \frac{m_i n_i}{1 + n_i \theta} = r_{Y^2}(\theta) - \frac{r_Y(\theta)^2}{r_1(\theta)}.$$

Solving the second equation in (15) with $\mu = \hat{\mu}(\theta)$ for κ thus gives

$$(24) \quad \hat{\kappa}(\theta) = \frac{N}{W + r_{Y^2}(\theta) + r_{B/m}(\theta) - \frac{r_Y(\theta)^2}{r_1(\theta)}}$$

$$(25) \quad = \frac{Nr_1(\theta)}{Wr_1(\theta) + r_{Y^2}(\theta)r_1(\theta) + r_1(\theta)r_{B/m}(\theta) - r_Y(\theta)^2}.$$

Substituting $\hat{\mu}(\theta)$ and $\hat{\kappa}(\theta)$ into the third and last equation in (16), we obtain the univariate rational equation

$$(26) \quad s_{Y^2}(\theta) - 2 \frac{r_Y(\theta)}{r_1(\theta)} s_Y(\theta) + \frac{r_Y(\theta)^2}{r_1(\theta)^2} s_1(\theta) + s_{B/m}(\theta) - \frac{r_1(\theta)}{\hat{\kappa}(\theta)} = 0,$$

where we have divided by the non-zero rational expression $\hat{\kappa}(\theta)$. According to (24), this is

$$(27) \quad s_{Y^2}(\theta) - 2 \frac{r_Y(\theta)}{r_1(\theta)} s_Y(\theta) + \frac{r_Y(\theta)^2}{r_1(\theta)^2} s_1(\theta) + s_{B/m}(\theta) - \frac{Wr_1(\theta) + r_{Y^2}(\theta)r_1(\theta) + r_1(\theta)r_{B/m} - r_Y(\theta)^2}{N} = 0.$$

Reexpress (26) in terms of the f and g polynomials as

$$(28) \quad \frac{g_{Y^2}(\theta)}{d(\theta)^2} - 2 \frac{f_Y(\theta)}{f_1(\theta)} \frac{g_Y(\theta)}{d(\theta)^2} + \frac{f_Y(\theta)^2}{f_1(\theta)^2} \frac{g_1(\theta)}{d(\theta)^2} + \frac{g_{B/m}}{d(\theta)^2} - \frac{W f_1(\theta) d(\theta) + f_{Y^2}(\theta) f_1(\theta) + f_1(\theta) f_{B/m} - f_Y(\theta)^2}{N d(\theta)^2} = 0.$$

The claim now follows by forming a common denominator. \square

The denominator given in (21) in Lemma 1 has degree $2M + 2(M - 1) = 4M - 2$. The numerator in (21) has degree $3(M - 1) + M = 4M - 3$; the highest degree term involves the within-group sum of squares W . The next two lemmas imply that, after cancelling common factors, the numerator of the univariate rational function from Lemma 1 has the degree claimed in the ML degree formula from Theorem 1.

Lemma 2. *If $m_t = 1$, then $(1 + n_t\theta)$ divides the numerator of the rational equation (21). Hence, the polynomial $d_1(\theta)$ of degree $M - M_2$ divides this numerator.*

Lemma 3. *If $d_1(\theta)$ is cleared from both the numerator and the denominator of the rational function given in (21), then the new numerator and denominator are relatively prime for generic sufficient statistics $\bar{Y}_1, \dots, \bar{Y}_M$, W , and B_j with $m_j \geq 2$.*

Proof of Lemma 2. Let $m_t = 1$. To show that $(1 + n_t\theta)$ divides the numerator, it is sufficient to show $(1 + n_t\theta)$ divides the sum of

$$(29) \quad N [f_1(\theta)^2 g_{Y^2}(\theta) - 2f_Y(\theta) f_1(\theta) g_Y(\theta) + f_Y(\theta)^2 g_1(\theta) + f_1(\theta)^2 g_{B/m}(\theta)]$$

and

$$(30) \quad -f_1(\theta)^2 [f_{Y^2}(\theta) f_1(\theta) - f_Y(\theta)^2 + f_1(\theta) f_{B/m}(\theta)].$$

The product $f_1(\theta)^2 g_{Y^2}(\theta)$ in the first term of (29) may be rewritten as

$$\begin{aligned} & \left[\sum_{i=1}^M m_i n_i \prod_{j \neq i} (1 + n_j \theta) \right] \left[\sum_{k=1}^M m_k n_k \prod_{l \neq k} (1 + n_l \theta) \right] \left[\sum_{r=1}^M m_r n_r^2 \bar{Y}_r^2 \prod_{s \neq r} (1 + n_s \theta)^2 \right] \\ &= \sum_{i=1}^M \sum_{k=1}^M \sum_{r=1}^M m_i m_k m_r n_i n_k n_r^2 \bar{Y}_r^2 \prod_{j \neq i} (1 + n_j \theta) \prod_{l \neq k} (1 + n_l \theta) \prod_{s \neq r} (1 + n_s \theta)^2. \end{aligned}$$

Combining this expression with the analogous expansions of the other three terms shows that the polynomial in (29) is equal to N times

$$(31) \quad \sum_{i=1}^M \sum_{k=1}^M \sum_{r=1}^M \left[(m_r \bar{Y}_r^2 - 2m_r \bar{Y}_i \bar{Y}_r + m_r \bar{Y}_i \bar{Y}_k + B_r) m_i m_k n_i n_k n_r^2 \right. \\ \left. \times \prod_{j \neq i} (1 + n_j \theta) \prod_{l \neq k} (1 + n_l \theta) \prod_{s \neq r} (1 + n_s \theta)^2 \right].$$

The polynomial in (30) can be expanded similarly. We find

$$(32) \quad f_{Y^2}(\theta) f_1(\theta) - f_Y(\theta)^2 + f_1(\theta) f_{B/m}(\theta) \\ = \sum_{i=1}^M \sum_{k=1}^M (m_k \bar{Y}_i^2 - m_k \bar{Y}_i \bar{Y}_k + B_k) m_i n_i n_k \prod_{j \neq i} (1 + n_j \theta) \prod_{l \neq k} (1 + n_l \theta).$$

Expanding $f_1(\theta)^2$ as well, we obtain that the polynomial in (30) is equal to

$$(33) \quad - \sum_{i=1}^M \sum_{k=1}^M \sum_{r=1}^M \sum_{u=1}^M \left[(m_k \bar{Y}_i^2 - m_k \bar{Y}_i \bar{Y}_k + B_k) m_i m_r m_u n_i n_k n_r n_u \right. \\ \left. \prod_{j \neq i} (1 + n_j \theta) \prod_{l \neq k} (1 + n_l \theta) \prod_{s \neq r} (1 + n_s \theta) \prod_{v \neq u} (1 + n_v \theta) \right].$$

Now notice that $(1 + n_t \theta)$ divides every summand in (31) and (33) unless $i = k = r = t$ in the first summation, or $i = k = r = u = t$ in the second summation. So it suffices to only consider these ‘diagonal’ terms. However, under the equality of indices, the quadratic expressions in the averages \bar{Y}_i cancel. Hence, the terms missing a factor of $(1 + n_t \theta)$ in (29) and (30) sum to

$$(34) \quad B_t n_t m_t^2 n_t^4 (N - m_t) \prod_{j \neq t} (1 + n_j \theta)^4.$$

Throughout the paper, we assume that we have at least two groups with at least one group size $n_i \geq 2$. Moreover, for generic data, $B_t = 0$ if and only if $m_t = 1$. Hence, for generic data, the expression in (34) is zero if and only if $m_t = 1$. We conclude that $d_1(\theta)$ divides the numerator of the rational function in (21). \square

Note that the last part of the above proof shows not only that $d_1(\theta)$ divides the numerator of (21), but that $(1 + n_t \theta)$ does not divide the numerator when $B_t \neq 0$, which holds generically if $m_t \geq 2$.

Proof of Lemma 3. Clearing $d_1(\theta)$ from the denominator in (21) yields the polynomial $N d_2(\theta) d(\theta) f_1(\theta)^2$. From the preceding comment, we know that $d_2(\theta)$ and the numerator are relatively prime for generic data $\bar{Y}_1, \dots, \bar{Y}_M$, $W > 0$, and $B_j > 0$ with $m_j \geq 2$. To establish our claim, we will first show that $f_1(\theta)$ does not share a common factor with the numerator by showing $f_1(\theta)$ and $f_Y(\theta)^2 g_1(\theta)$ to be relatively prime; all terms other than $f_Y(\theta)^2 g_1(\theta)$ in the numerator of (21) are multiples of $f_1(\theta)$. Then, we will show that after clearing $d_1(\theta)$ in (21), $d_1(\theta)$ and the new numerator are relatively prime.

Let $\theta_1, \dots, \theta_{M-1}$ be the (possibly complex) roots of the degree $M-1$ polynomial $f_1(\theta)$. For each $1 \leq k \leq M-1$, consider the linear form $f_Y(\theta_k)$ in the polynomial ring $\mathbb{C}[\bar{Y}_1, \dots, \bar{Y}_M]$. Let $V(f_Y(\theta_k)) \subset \mathbb{C}^M$ be the zero locus of $f_Y(\theta_k)$. Each set $V(f_Y(\theta_k))$ is a hyperplane of dimension $M-1$. Thus, the union $\cup_{k=1}^{M-1} V(f_Y(\theta_k))$ is an $M-1$ dimensional algebraic subset of \mathbb{C}^M . A generic vector of group means $(\bar{Y}_1, \dots, \bar{Y}_M)$ lies outside this lower-dimensional set, which means that $f_1(\theta)$ and $f_Y(\theta)$ are relatively prime for generic data.

To show that $f_1(\theta)$ and $g_1(\theta)$ are relatively prime, assume $\theta_0 = a + ib$ is a root of $f_1(\theta)$ and $g_1(\theta)$. Since $g_1(\theta)$ is a sum of squares that is positive on \mathbb{R} , we must have $\theta_0 \notin \mathbb{R}$ and hence $b \neq 0$. Without loss of generality, let n_1 be the least of the group sizes n_i . Rewriting $f_1(\theta_0) = 0$, we get

$$(35) \quad n_1 = - \frac{\sum_{i=2}^M m_i n_i \prod_{j \neq i} (1 + n_j \theta_0)}{m_1 \prod_{j \neq 1} (1 + n_j \theta_0)} = - \sum_{i=2}^M \frac{m_i n_i (1 + n_1 \theta_0)}{m_1 (1 + n_i \theta_0)}.$$

The imaginary part of the right side of this equation must equal 0 since n_1 is an integer. Substituting $a + ib$ for θ_0 , the imaginary part of (35) is

$$b \sum_{i=2}^M \left(\frac{m_i n_i}{m_1} \right) \frac{(n_i - n_1)}{(1 + n_i a)^2 + (n_i b)^2}.$$

Since each term in the sum is positive, we obtain that $b = 0$. Consequently, $\theta_0 \in \mathbb{R}$, which is a contradiction. Therefore, $f_1(\theta)$ and $g_1(\theta)$ are relatively prime.

It remains to show that the numerator and denominator obtained by clearing the factor $d_1(\theta)$ in (21) are relatively prime for generic data. We claim that if $m_t = 1$ then $(1 + n_t \theta)$ divides

$$(36) \quad \frac{f_1(\theta)^2 g_{Y^2}(\theta) - 2f_Y(\theta) f_1(\theta) g_Y(\theta) + f_Y(\theta)^2 g_1(\theta) f_1(\theta)^2 g_{B/m}(\theta)}{d_1(\theta)},$$

while $d_1(\theta)$ and

$$(37) \quad W f_1(\theta) d_2(\theta) + \frac{f_{Y^2}(\theta) f_1(\theta) - f_Y(\theta)^2 + f_1(\theta) f_{B/m}(\theta)}{d_1(\theta)} =: W f_1(\theta) d_2(\theta) + F(\theta)$$

are relatively prime for generic data.

The ratio in (36) equals (31) divided by $d_1(\theta)$. We may rewrite (31) as

$$(38) \quad \sum_{i=1}^M \sum_{k=1}^M \sum_{r=1}^M \left[(m_r \bar{Y}_r^2 - m_r \bar{Y}_i \bar{Y}_r - m_r \bar{Y}_k \bar{Y}_r + m_r \bar{Y}_i \bar{Y}_k + B_r) m_i m_k n_i n_k n_r^2 \right. \\ \left. \times \prod_{j \neq i} (1 + n_j \theta) \prod_{l \neq k} (1 + n_l \theta) \prod_{s \neq r} (1 + n_s \theta)^2 \right].$$

It is clear that the square $(1 + n_t \theta)^2$ divides all terms in the sum (38) except those for $r = i = t$ or $r = k = t$. However, the quadratic form in the averages \bar{Y}_i vanishes if $r = i$ or $r = k$. Since the terms in question have $r = t$, and $B_r = B_t = 0$ because $m_t = 1$, we conclude that $(1 + n_t \theta)^2$ divides the entire sum (38), which proves that $d_1(\theta)$ divides the ratio in (36).

We are left to show that $d_1(\theta)$ and $W f_1(\theta) d_2(\theta) + F(\theta)$ are relatively prime for generic data. Let $\theta_1, \dots, \theta_{M-M_2}$ be the roots of $d_1(\theta)$; each root is equal to $-1/n_i$ for some index i . Since the n_i are distinct, no root of $d_1(\theta)$ is a root of $d_2(\theta)$. Moreover, it is easy to see that no root of $d_1(\theta)$ is a root of $f_1(\theta)$. Now let I be the ideal generated by the $M - M_2$ polynomials $W f_1(\theta_k) d_2(\theta_k) + F(\theta_k)$ in the polynomial ring $\mathbb{C}[W, \bar{Y}_1, \dots, \bar{Y}_M, B'_1, \dots, B'_{M_2}]$, where the B'_i stand for the between-group sums of squares B_i with multiplicity $m_i \geq 2$. Pick sufficient statistics $W = \bar{Y}_1 = \dots = \bar{Y}_M \neq 0$ and $B'_1 = \dots = B'_M = 0$. Since no root of $d_1(\theta)$ is a root of $d_2(\theta)$ or $f_1(\theta)$, (32) implies that for these special data $W f_1(\theta_k) d_2(\theta_k) + F(\theta_k) \neq 0$ for each k . The zero locus $V(I)$ is thus a proper algebraic subset of \mathbb{C}^{M+M_2+1} . Such a set is of lower dimension and, thus, $d_1(\theta)$ and $W f_1(\theta) d_2(\theta) + F(\theta)$ are relatively prime for generic data. \square

4. PROOF OF FORMULA FOR REML DEGREE

For the proof of the REML degree formula in Theorem 1, we proceed in the same way as for the ML degree. We begin by deriving the univariate rational function whose number of roots is the REML degree.

Lemma 4. *Consider the rational function whose numerator is*

$$(39) \quad (g_1(\theta) - f_1(\theta)^2)[Wf_1(\theta)d(\theta) + f_{Y^2}(\theta)f_1(\theta) - f_Y(\theta)^2 + f_1(\theta)f_{B/m}] + \\ (N-1)[f_1(\theta)^2g_{Y^2}(\theta) - 2f_Y(\theta)f_1(\theta)g_Y(\theta) + f_Y(\theta)^2g_1(\theta) + f_1(\theta)^2g_{B/m}(\theta)]$$

and denominator is

$$(40) \quad d(\theta)f_1(\theta)[Wf_1(\theta)d(\theta) + f_{Y^2}(\theta)f_1(\theta) - f_Y(\theta)^2 + f_1(\theta)f_{B/m}].$$

The REML degree is the degree of the numerator of this rational function after clearing common factors from the given numerator and denominator.

Proof. The equation $\partial\bar{\ell}/\partial\kappa = 0$ has the unique solution

$$\hat{\kappa}(\theta) = \frac{N-1}{W + \sum_{i=1}^M \frac{m_i n_i}{1+n_i\theta} (\bar{Y}_i - \hat{\mu}(\theta))^2 + \sum_{i=1}^M \frac{n_i}{1+n_i\theta} B_i};$$

compare (19). Substituting $\hat{\kappa}(\theta)$ into the partial derivative $\partial\bar{\ell}/\partial\theta$ yields the univariate function

$$(41) \quad -\sum_{i=1}^M \frac{m_i n_i}{1+n_i\theta} + \frac{\sum_{i=1}^M \frac{m_i n_i^2}{(1+n_i\theta)^2}}{\sum_{i=1}^M \frac{m_i n_i}{1+n_i\theta}} \\ + \hat{\kappa}(\theta) \left[\sum_{i=1}^M \frac{m_i n_i^2}{(1+n_i\theta)^2} (\bar{Y}_i - \hat{\mu}(\theta))^2 + \sum_{i=1}^M \frac{n_i^2}{(1+n_i\theta)^2} B_i \right] = 0;$$

recall (20). We can now simplify and rewrite (41), forming a common denominator, to obtain the desired rational function. \square

The degree of the numerator in Lemma 4 is $4M-3$, but it shares common factors with the denominator. In fact, in the proof of Lemma 2, we have shown that $d_1(\theta)$ divides $f_{Y^2}(\theta)f_1(\theta) - f_Y(\theta)^2 + f_1(\theta)f_{B/m}$. Thus, $d_1(\theta)^2$ divides the denominator from (40). To prove Theorem 1, it remains to prove the following two facts.

Lemma 5. *The polynomial $d_1(\theta)^2$ divides the numerator (39).*

Lemma 6. *After clearing $d_1(\theta)^2$ from (39) and (40), the new numerator and new denominator are relatively prime for generic data.*

Proof of Lemma 5. From the proof of Lemma 2, we know that $d_1(\theta)$ divides the polynomial $f_{Y^2}(\theta)f_1(\theta) - f_Y(\theta)^2 + f_1(\theta)f_{B/m}$. Moreover, as shown in the proof of Lemma 3, the square $d_1(\theta)^2$ divides

$$f_1(\theta)^2g_{Y^2}(\theta) - 2f_1(\theta)f_Y(\theta)g_Y(\theta) + f_Y(\theta)^2g_1(\theta) + f_1(\theta)^2g_{B/m}(\theta).$$

To complete the proof of the present lemma, it suffices to show that $d_1(\theta)$ divides $g_1(\theta) - f_1(\theta)^2$. However, with some distributing and grouping, we see

$$\begin{aligned} g_1(\theta) - f_1(\theta)^2 &= \\ &= \sum_{i=1}^M m_i n_i^2 \prod_{j \neq i} (1 + n_j \theta)^2 - \sum_{i=1}^M \sum_{k=1}^M m_i m_k n_i n_k \prod_{j \neq i} (1 + n_j \theta) \prod_{l \neq k} (1 + n_l \theta) \\ &= \sum_{i=1}^M (m_i - m_i^2) \prod_{j \neq i} (1 + n_j \theta) - \sum_{i=1}^M \sum_{k > i}^M 2n_i n_k \prod_{j \neq i} (1 + n_j \theta) \prod_{l \neq k} (1 + n_l \theta), \end{aligned}$$

which is divisible by $(1 + n_t \theta)$ if and only if $m_t = 1$. \square

Proof of Lemma 6. We first show that if $m_t \geq 2$, then, for generic data, $(1 + n_t \theta)$ and the numerator from (39) are relatively prime. Consider

$$(42) \quad (g_1(\theta) - f_1(\theta)^2) [f_{Y^2}(\theta) f_1(\theta) - f_Y(\theta)^2 + f_1(\theta) f_{B/m}(\theta)] + (N-1) [f_1(\theta)^2 g_{Y^2}(\theta) - 2f_Y(\theta) f_1(\theta) g_Y(\theta) + f_Y(\theta)^2 g_1(\theta) + g_{B/m}(\theta) f_1(\theta)^2].$$

Using the results from the proof of Lemma 2 and writing out the involved summations, (42) is seen to be equal to

$$\begin{aligned} (43) \quad & \left(\sum_{i=1}^M \sum_{k=1}^M \sum_{r=1}^M (m_k \bar{Y}_i^2 - m_k \bar{Y}_i \bar{Y}_k + B_k) m_i m_r n_i n_k n_r^2 \right. \\ & \quad \left. \prod_{j \neq i} (1 + n_j \theta) \prod_{l \neq k} (1 + n_l \theta) \prod_{s \neq r} (1 + n_s \theta)^2 \right) \\ & - \left(\sum_{i=1}^M \sum_{k=1}^M \sum_{r=1}^M \sum_{u=1}^M (m_k \bar{Y}_i^2 - m_k \bar{Y}_i \bar{Y}_k + B_k) m_i m_r m_u n_i n_k n_r n_u \right. \\ & \quad \left. \prod_{j \neq i} (1 + n_j \theta) \prod_{l \neq k} (1 + n_l \theta) \prod_{s \neq r} (1 + n_s \theta) \prod_{v \neq u} (1 + n_v \theta) \right) \\ & + (N-1) \left[\sum_{i=1}^M \sum_{k=1}^M \sum_{r=1}^M (m_r \bar{Y}_i^2 - 2m_r \bar{Y}_i \bar{Y}_r + m_r \bar{Y}_i \bar{Y}_k + B_r) m_i m_k n_i n_k n_r^2 \right. \\ & \quad \left. \prod_{j \neq i} (1 + n_j \theta) \prod_{l \neq k} (1 + n_l \theta) \prod_{s \neq r} (1 + n_s \theta)^2 \right]. \end{aligned}$$

The factor $(1 + n_t \theta)$ divides every summand in the above summations unless $t = i = k = r = u$, so it suffices to only consider these terms. Letting $t = i = k = r = u$, the terms missing a factor of $(1 + n_t \theta)$ sum to a term we already encountered, namely, that in (34). The discussion following display (34) shows that if the data is generic and $m_t \geq 2$, then $(1 + n_t \theta)$ does not divide the numerator given in (39).

Continuing to work through the factors of the denominator from (40), assume that θ_0 is a root of $f_1(\theta)$. Then everything vanishes in the numerator except for two terms $-g_1(\theta_0) f_Y(\theta_0)^2$ and $(N-1) f_Y(\theta_0)^2 g_1(\theta_0)$, which add to $(N-2) f_Y(\theta_0)^2 g_1(\theta_0)$. From the proof of Lemma 3, we know $f_Y(\theta_0)^2 g_1(\theta_0) \neq 0$ for generic data, so since

we are working under the assumption of at least two groups and at least one group size $n_i \geq 2$, the numerator and $f_1(\theta)$ are relatively prime for generic data.

Finally, we need to show

$$H(\theta) := f_1(\theta)^2 g_{Y^2}(\theta) - 2f_1(\theta)f_Y(\theta)g_Y(\theta) + f_Y(\theta)^2 g_1(\theta) + f_1(\theta)^2 g_{B/m}(\theta)$$

and

$$G(\theta) := Wf_1(\theta)d_2(\theta) + F(\theta),$$

are relatively prime for generic data $W, \bar{Y}_1, \dots, \bar{Y}_M$, and B_i with $m_i \geq 2$; the polynomial $F(\theta)$ was defined in (37). We will again denote the between-group sums of squares with multiplicities $m_i \geq 2$ as B'_1, \dots, B'_{M_2} . By a standard algebraic results, the polynomials $G(\theta)$ and $H(\theta)$ share a common root θ if and only if a certain polynomial in their coefficients vanishes; this polynomial is called the resultant and we denote it by $\text{Res}(G, H)$. Since both $H(\theta)$ and $G(\theta)$ have coefficients that are polynomials in the sufficient statistics $W, \bar{Y}_1, \dots, \bar{Y}_M$, and B'_1, \dots, B'_{M_2} , we may regard $\text{Res}(G, H)$ as a polynomial in the ring $\mathbb{C}[W, \bar{Y}_1, \dots, \bar{Y}_M, B'_1, \dots, B'_{M_2}]$. By Lemma 3, for any given generic choice of $\bar{Y}_1, \dots, \bar{Y}_M, B'_1, \dots, B'_{M_2}$, a root θ_0 of H is not a root of $f_1(\theta)$ or $d_2(\theta)$. Hence, θ_0 is a root of G if and only if

$$(44) \quad W = -\frac{F(\theta_0)}{d_2(\theta_0)f_1(\theta_0)}.$$

Picking W not to satisfy (44) shows that $\text{Res}(G, H)$ is not the zero polynomial in $\mathbb{C}[W, \bar{Y}_1, \dots, \bar{Y}_M, B'_1, \dots, B'_{M_2}]$. Hence, the zero locus of $\text{Res}(G, H)$ is a set of lower dimension, and we conclude that H and G are relatively prime for generic data. \square

5. GENERAL MEAN STRUCTURE IN THE ONE-WAY LAYOUT

The one-way layout as specified in (1) postulates a common mean μ for all observations Y_{ij} . Often the interest is instead in a more general mean space. Formally, consider the model

$$(45) \quad Y_{ij} = \mu_{ij} + \alpha_i + \varepsilon_{ij}, \quad i = 1, \dots, q, \quad j = 1, \dots, n_i,$$

where the random effects $\alpha_i \sim \mathcal{N}(0, \tau)$ and the error terms $\varepsilon_{ij} \sim \mathcal{N}(0, \omega)$ are again all mutually independent. However, the array of means (μ_{ij}) may now belong to a linear subspace of \mathbb{R}^N that we assumed to be spanned by the independent columns of a full rank design matrix $X \in \mathbb{R}^{N \times p}$; as before, $N = n_1 + \dots + n_q$ is the sample size. In other words,

$$(46) \quad \text{vec}(\mu_{ij}) = X\beta$$

for some unknown (fixed) mean parameter vector $\beta \in \mathbb{R}^p$.

ML and REML estimation with more general mean structure can be approached algebraically in the exactly the same way as before. It is convenient to reparametrize the covariance matrix in terms of $\kappa = 1/\omega$ and $\theta = \tau/\omega$. For known covariance parameters, the ML estimate $\hat{\beta}(\theta)$ of β is obtained by generalized least squares and depends on θ but not on κ . For fixed θ , it is then also straightforward to solve the ML or REML equations for κ . This way we may reduce algebraic solution of the likelihood equations to solving a single rational equation in θ . In this section we demonstrate that the involved algebraic computations are feasible in a larger example. Before going into the details of the example, we would like to offer the following conjecture based on numerical experiments with smaller models and randomly chosen design matrices. It states that the ML and REML degrees for the

model specified by (45) and (46) cannot exceed the largest possible respective degrees in the model with common mean μ . Recall that the largest degrees arise in the entirely unbalanced case with group sizes n_1, \dots, n_q that are pairwise distinct.

Conjecture 1. *For any design matrix $X \in \mathbb{R}^{N \times p}$ that has the vector $(1, \dots, 1)^T$ in its column span $\text{span}(X)$, the ML degree for the one-way layout with mean space $\text{span}(X)$ and q random group effects is bounded above by $3q - 3$. Similarly, the REML degree is bounded above by $2q - 3$.*

According to this conjecture, the degrees would grow only linearly with the number of groups, which would suggest that a moderately large number of unbalanced groups can be handled in algebraic computations.

Example 3. With the goal of providing linguistic support for an African origin of modern humans, Atkinson [Atk11] fits regression models to data on the phonemic diversity of languages. The data, which can be obtained from the journal's online supplementary material, concern $N = 504$ languages that are classified into $q = 109$ language families. Besides quantitative summary measures of phonemic diversity, the available information includes the size of the population speaking each language and the distance between a chosen center for each language and an inferred origin in Africa, the latter being the main covariate of interest.

One model of interest in this application is a one-way layout with groups corresponding to the language families. The response Y_{ij} is the phonemic diversity of the j th language in the i th family, which, as in (45) and (46), is modelled as

$$(47) \quad Y_{ij} = \beta_0 + \beta_1 \log(P_{ij}) + \beta_2 D_{ij} + \alpha_i + \epsilon_{ij}, \quad i = 1, \dots, q, \quad j = 1, \dots, n_i.$$

Here, P_{ij} stands for the population size and D_{ij} is the distance from the origin in Africa. As can be expected, the data is unbalanced. The group sizes n_1, \dots, n_{109} fall into the range from 1 to 62. There are $M = 17$ distinct group sizes of which $M_2 = 9$ have multiplicity two or larger. Hence, by Theorem 1, the one-way layout with all means equal has ML degree 57 and REML degree 49. However, as we show next, the mean structure can affect the ML and REML degree.

Computations we did using the software Maple show that the ML degree of the model given by (47) is 83, whereas the REML degree is 71. Exact computations in analogy to the ones given in Example 1 produce large integer coefficients, too large to display on paper but easily handled by a computer. Solving the polynomial equations for ML and REML numerically, each equation is seen to have a unique positive root, namely,

$$(48) \quad \hat{\theta}_{\text{ML}} \approx 0.3706 \quad \text{and} \quad \hat{\theta}_{\text{REML}} \approx 0.3853.$$

Each root gives a local and, thus, global maximum of the concerned likelihood function. We remark that the ML equation has twelve negative real roots. The REML equation has no other real roots. Running the numerical optimizers implemented in the R package `lme4` yields estimates that agree with (48) in all the given digits. As in Example 1, the fact that our two univariate polynomials each have a unique positive root manifests itself in a single sign change in the coefficient sequence. Finally, we remark that when omitting either the covariate $\log(P)$ or the covariate D , the ML degree drops to 72 and the REML degree drops to 61.

6. BALANCED TWO-WAY LAYOUTS

Suppose we have observations Y_{ijk} that are cross-classified according to two factors and model the observations in an additive two-way layout as

$$(49) \quad Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk},$$

$$i = 1, \dots, r, \quad j = 1, \dots, q, \quad k = 1, \dots, n.$$

The terms $\alpha_i \sim \mathcal{N}(0, \tau_1)$ and $\beta_j \sim \mathcal{N}(0, \tau_2)$ are normally distributed random effects. The error terms are distributed as $\epsilon_{ijk} \sim \mathcal{N}(0, \omega)$, and all the random variables α_i , β_j and ϵ_{ijk} are mutually independent. Finally, there is one (fixed) mean parameter $\mu \in \mathbb{R}$. A related model is obtained by including random interaction terms $\gamma_{ij} \sim \mathcal{N}(0, \tau_{12})$ in the defining equations

$$(50) \quad Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk},$$

$$i = 1, \dots, r, \quad j = 1, \dots, q, \quad k = 1, \dots, n.$$

The interaction terms γ_{ij} are again mutually independent and independent of all other random variables appearing on the right hand side of (50).

The models in (49) and (50) are balanced; the groups of observations Y_{ij1}, \dots, Y_{ijn} specified by the different index pairs (i, j) are all of size n . It is known that REML leads to closed form estimates for each of the two balanced models; compare [Hoc85, SCM92, SO04]. In other words, the REML degree of either model is one. ML estimation, however, presents a non-trivial algebraic problem. The ML degree can be derived using Gröbner basis calculations, and we see that balanced two-way layouts have closed form ML estimates in the sense of Cardano's formula.

Theorem 2. *The ML degree of balanced additive two-way layout with random effects is four. The same holds for the model with random interaction.*

Proof. Define the sum of squares

$$SSA = \sum_{i=1}^r qn(\bar{Y}_{i..} - \bar{Y}_{...})^2,$$

$$SSB = \sum_{j=1}^q rn(\bar{Y}_{.j.} - \bar{Y}_{...})^2,$$

$$SSAB = \sum_{i=1}^r \sum_{j=1}^q n(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2,$$

$$SSE = \sum_{i=1}^r \sum_{j=1}^q \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{ij.})^2,$$

where we use the convention that the overbar indicates that an average was formed and the ‘.’ subscripts specify which indices were averaged over.

(No interaction) The ML equations for the additive model given by (49) are derived, for instance, in Chapter 4.7.d of [SCM92] and in Chapter 3 of [SO04]. One equation leads to the ML estimator

$$\hat{\mu} = \bar{Y}_{...}.$$

The rational equations for the variance components may be written as

$$(51) \quad \frac{rqn - r - q + 1}{\omega} - \frac{1}{\omega + qn\tau_1 + rn\tau_2} = \frac{SSAB + SSE}{\omega^2},$$

$$(52) \quad \frac{r - 1}{\omega + qn\tau_1} + \frac{1}{\omega + qn\tau_1 + rn\tau_2} = \frac{SSA}{(\omega + qn\tau_1)^2},$$

$$(53) \quad \frac{q - 1}{\omega + rn\tau_2} + \frac{1}{\omega + qn\tau_1 + rn\tau_2} = \frac{SSB}{(\omega + rn\tau_2)^2}.$$

Clearing the common denominators ω^2 , $(\omega + qn\tau_1)^2$, $(\omega + rn\tau_2)^2$, and $\omega + qn\tau_1 + rn\tau_2$ gives a polynomial equation system. However, multiplying each equation with the relevant product of these denominators introduces new solutions that are not solutions of the original rational equations. Using saturation as explained in Chapter 2 of [DSS09], we can remove these extraneous solutions and obtain a polynomial equation system of degree 4. (We remark that software such as Maple is able to produce a lexicographic Gröbner basis over the field of fractions in r , q , n , and the four sums of squares.)

(*With interaction*) Chapter 4.7.d of [SCM92] also gives the ML equations for the model with interaction defined by (50); see also Chapter 4 of [SO04]. Two equations determine the ML estimators

$$\hat{\mu} = \bar{Y}_{...}, \quad \hat{\omega} = \frac{SSE}{rq(n-1)}.$$

The rational equations for the remaining variance components can be written as

$$(54) \quad \frac{(r-1)(q-1)}{\hat{\omega} + n\tau_{12}} - \frac{1}{\hat{\omega} + qn\tau_1 + rn\tau_2 + n\tau_{12}} = \frac{SSAB}{(\hat{\omega} + n\tau_{12})^2},$$

$$(55) \quad \frac{r-1}{\hat{\omega} + qn\tau_1 + n\tau_{12}} + \frac{1}{\hat{\omega} + qn\tau_1 + rn\tau_2 + n\tau_{12}} = \frac{SSA}{(\hat{\omega} + qn\tau_1 + n\tau_{12})^2},$$

$$(56) \quad \frac{q-1}{\hat{\omega} + rn\tau_2 + n\tau_{12}} + \frac{1}{\hat{\omega} + qn\tau_1 + rn\tau_2 + n\tau_{12}} = \frac{SSB}{(\hat{\omega} + rn\tau_2 + n\tau_{12})^2}.$$

Clearing the denominators carefully via saturation yields a polynomial equation system of degree 4. (Again, a lexicographic Gröbner basis can be obtained with r , q , n and the sums of squares as parameters to the equations.) \square

We briefly illustrate algebraic computation of the ML estimators in an example that involves the additive two-way layout.

Example 4. The R package `lme4` contains data from experiments for an assessment of the variability between samples of penicillin. The data are described in detail in [DG72]. The response is a diameter measurement of the zone in which growth of an organism is inhibited by the penicillin. The experiments are cross-classified according to the assay plate and the penicillin sample used. The former is a factor with $r = 24$ levels, the latter has $q = 6$ levels. There are no replications to be considered in this case, that is, $n = 1$. We will consider the additive model for which the relevant sums of squares are

$$SSA = 105\frac{8}{9}, \quad SSB = 449\frac{2}{9}, \quad SSAB + SSE = 34\frac{7}{9}.$$

Using the saturation computation alluded to in the proof of Theorem 2, we obtain the polynomial equation system

$$\begin{aligned}
204808595904 \omega^4 - 1801205257140 \omega^3 + 2545119731943 \omega^2 - 1070402996440 \omega \\
+ 139045932165 &= 0, \\
2481278604010272 \tau_1 + 507582172417738176 \omega^3 - 4309720916424828084 \omega^2 \\
+ 4998133978544934251 \omega - 1133204709683307975 &= 0, \\
2481278604010272 \tau_2 + 534435082556924736 \omega^3 - 4538697213124439100 \omega^2 \\
+ 5270402449572117709 \omega - 1201351121037374475 &= 0.
\end{aligned}$$

This polynomial system has the same solution set as the original rational ML equations. The polynomials on the left hand sides of the equations form a lexicographic Gröbner basis and are readily solved. First, solve the quartic equation in ω . Next, plug each of the four solutions for ω into the other two equations and solve the resulting linear equations for τ_1 and τ_2 , respectively. In the present example, all four solutions are real but only one is feasible with $\omega, \tau_1, \tau_2 \geq 0$. This solution is

$$\hat{\omega} = 0.302425, \quad \hat{\tau}_1 = 0.714992, \quad \hat{\tau}_2 = 3.135188.$$

It defines the unique global maximum of the likelihood function.

7. CONCLUSION

This paper takes a first step towards understanding the algebraic complexity of ML and REML estimation in linear mixed models. Our main results in Theorem 1 concern the unbalanced one-way layout with common mean for all observations. It would be interesting to generalize the results to one-way classifications with more complicated mean spaces; recall Conjecture 1. Similarly, it would be interesting to study unbalanced two- and higher-way layouts, although these models would require more sophisticated mathematical treatment because it is no longer possible to analyze a single univariate rational equation; compare Section 6.

A remarkable feature common to Examples 1 and 3 is that Descartes' rule of sign applied to a univariate polynomial in the variance ratio θ reveals that there is a unique feasible solution to the ML/REML equations. The same was true for many other examples of unbalanced one-way classifications that we computed. This said, we also saw cases with more than one sign change and the number of positive solutions for θ not matching up with the sign changes.

To our knowledge, the literature does not supply many examples of linear mixed models with multimodal likelihood functions. We conclude by giving two simulated examples that demonstrate the mathematical possibility of more than one mode. Such examples were rare in our simulations, which is in agreement with findings of [SM84] who also treat the unbalanced one-way layout. While uniqueness of local optima is not explicitly discussed in [SM84], the authors remark in their conclusion that “varying the iteration starting point slightly affects the rate of convergence, but not the [mean square errors] or biases of the [ML and REML] estimators.” The examples we give involve three positive roots to the ML or REML equations for the variance ratio θ . We do not know of examples with more positive roots.

Example 5. Consider the one-way layout with a single grand mean μ from (1). Take $q = 5$ groups of sizes

$$n_1 = 2, \quad n_2 = 5, \quad n_3 = 10, \quad n_4 = 20, \quad n_5 = 50.$$

Let the sufficient statistics be the five group averages

$$\begin{aligned} \bar{Y}_1 &= -\frac{73571}{14273} \approx -5.1546, & \bar{Y}_2 &= \frac{13781}{78326} \approx 0.1759, \\ \bar{Y}_3 &= -\frac{13277}{92152} \approx -0.1441, & \bar{Y}_4 &= \frac{31207}{202567} \approx 0.1541, \\ \bar{Y}_5 &= -\frac{15713}{24121} \approx -0.6514, \end{aligned}$$

and the within-group sum of squares

$$W = \frac{116487}{421} \approx 276.69.$$

The univariate ML equation in θ has three nonnegative solutions, namely,

$$\hat{\theta}_{\text{ML},1} \approx 0.00838738, \quad \hat{\theta}_{\text{ML},2} \approx 0.118458, \quad \hat{\theta}_{\text{ML},3} \approx 0.338944;$$

having specified six digits we should add that the solutions were computed treating the above rational fractions as the input. The solution $\hat{\theta}_{\text{ML},1}$ yields the global maximum of the likelihood function, whereas $\hat{\theta}_{\text{ML},2}$ and $\hat{\theta}_{\text{ML},3}$ determine a saddle point and local maximum, respectively. In contrast, the restricted likelihood function has a unique local and global maximum for

$$\hat{\theta}_{\text{REML}} \approx 0.771763.$$

The data was simulated from the model with mean $\mu_0 = 0$, and variance components $\tau_0 = 3$ and $\omega_0 = 2$, which gives $\theta_0 = 3/2$.

Example 6. Continuing with the setup from Example 5, change the sufficient statistics to

$$\begin{aligned} \bar{Y}_1 &= \frac{230081}{40206} \approx 5.7226, & \bar{Y}_2 &= \frac{721282}{5630371} \approx 0.1281, \\ \bar{Y}_3 &= \frac{29305}{95646} \approx 0.3064, & \bar{Y}_4 &= \frac{15365}{37988} \approx 0.4045, \\ \bar{Y}_5 &= -\frac{569}{40932} \approx -0.0139, \end{aligned}$$

and

$$W = \frac{755002}{1759} \approx 429.22.$$

Now, all real solutions to the ML equations are negative. Thus, the global maximum of the likelihood is achieved at the boundary point $\hat{\theta}_{\text{ML}} = 0$. In contrast, the REML equations have three feasible solutions for θ , namely,

$$\hat{\theta}_{\text{REML},1} \approx 0.00492193, \quad \hat{\theta}_{\text{REML},2} \approx 0.159465, \quad \hat{\theta}_{\text{REML},3} \approx 0.2414611.$$

The solution $\hat{\theta}_{\text{REML},1}$ gives the global maximum of the restricted likelihood function. The solutions $\hat{\theta}_{\text{REML},2}$ and $\hat{\theta}_{\text{REML},3}$ determine a saddle point and a local maximum, respectively. The data was simulated as in Example 5.

Readers experimenting with the two examples just given will find the likelihood functions to be rather flat between the three stationary points, which give log-likelihood values that differ by less than 0.1.

In both Example 5 and Example 6, the first group is of the smallest size but has group mean that is largest in absolute value. The other means are comparatively close to each other. We experimented with permuting the means, while holding the group sizes fixed. In Example 6, eight out of 120 permutations give bimodal

restricted likelihood functions. Two permutations yield three positive roots to the REML equations. The other six cases have two positive roots, and one of the two local maxima occurs for $\theta = 0$. The eight permutations generate the group of permutations that keep the first mean fixed. In this example, there is clearly negative correlation between the group sizes n_i and the group means \bar{Y}_i . (In practice, such dependence could arise from selection effects.) The eight permutations of interest turn out to give the eight most negative correlations between group sizes and means. In similar experiments for Example 5, which features positive correlation between group sizes and means, bimodal likelihood functions are obtained for 18 permutations. Again, these permutations keep the first mean fixed. Only three permutations give three positive roots to the ML equations. The 18 permutations include the top six permutations in terms of large positive correlation but also the permutation whose associated correlation ranks 43rd.

While dependence between group means and sizes plays a role in Examples 5 and 6, the precise interplay between them appears to be subtle. For instance, when varying the mean \bar{Y}_1 in Example 5 and keeping all other sufficient statistics fixed, we find that there are three positive roots to the ML equations when $-5.47 \leq \bar{Y}_1 \leq -5.08$ but a unique root otherwise; we experimented with a grid of values in $[-10, 10]$. In particular, the likelihood function is unimodal for larger negative values of \bar{Y}_1 . It would be interesting, but presumably difficult, to get a better understanding of the semi-algebraic set of sufficient statistics that give (restricted) likelihood functions with more than one local maximum.

REFERENCES

- [Atk11] Quentin D. Atkinson, *Phonemic diversity supports a serial founder effect model of language expansion from Africa*, *Science* **332** (2011), no. 6027, 346–349.
- [BHR07] Max-Louis G. Buot, Serkan Hoşten, and Donald St. P. Richards, *Counting and locating the solutions of polynomial systems of maximum likelihood equations. II. The Behrens-Fisher problem*, *Statist. Sinica* **17** (2007), no. 4, 1343–1354. MR 2398599 (2009c:62057)
- [CHKS06] Fabrizio Catanese, Serkan Hoşten, Amit Khetan, and Bernd Sturmfels, *The maximum likelihood degree*, *Amer. J. Math.* **128** (2006), no. 3, 671–697. MR 2230921 (2007m:13036)
- [DG72] Owen L. Davies and Peter L. Goldsmith (eds.), *Statistical methods in research and production*, 4th ed., Hafner, 1972.
- [DM99] Eugene Demidenko and Hélène Massam, *On the existence of the maximum likelihood estimate in variance components models*, *Sankhyā Ser. A* **61** (1999), no. 3, 431–443. MR 1743550
- [DSS09] Mathias Drton, Bernd Sturmfels, and Seth Sullivant, *Lectures on algebraic statistics*, Birkhäuser Verlag AG, Basel, Switzerland, 2009.
- [Far06] Julian J. Faraway, *Extending the linear model with R*, Texts in Statistical Science Series, Chapman & Hall/CRC, Boca Raton, FL, 2006, Generalized linear, mixed effects and nonparametric regression models. MR 2192856
- [HKS05] Serkan Hoşten, Amit Khetan, and Bernd Sturmfels, *Solving the likelihood equations*, *Found. Comput. Math.* **5** (2005), no. 4, 389–407. MR 2189544
- [Hoc85] R. R. Hocking, *The analysis of linear models*, Brooks/Cole Publishing Co., Monterey, CA, 1985. MR 805982 (86k:62120)
- [HS10] Serkan Hoşten and Seth Sullivant, *The algebraic complexity of maximum likelihood estimation for bivariate missing data*, Algebraic and geometric methods in statistics, Cambridge Univ. Press, Cambridge, 2010, pp. 123–133. MR 2642662
- [Jia07] Jiming Jiang, *Linear and generalized linear mixed models and their applications*, Springer Series in Statistics, Springer, New York, 2007. MR 2308058 (2007m:62002)
- [MN89] P. McCullagh and J. A. Nelder, *Generalized linear models*, 2nd ed., Monographs on Statistics and Applied Probability, Chapman & Hall, London, 1989.

- [SCM92] Shayle R. Searle, George Casella, and Charles E. McCulloch, *Variance components*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons Inc., New York, 1992, A Wiley-Interscience Publication. MR 1190470 (93m:62054)
- [SM84] William H. Swallow and John F. Monahan, *Monte Carlo comparison of ANOVA, MIVQUE, REML, and ML estimators of variance components*, *Technometrics* **26** (1984), no. 1, 47–57.
- [SO04] Hardeo Sahai and Mario Miguel Ojeda, *Analysis of variance for random models. Vol. I. Balanced data*, Birkhäuser Boston Inc., Boston, MA, 2004, Theory, methods, applications and data analysis. MR 2374911 (2008k:62006)
- [SO05] ———, *Analysis of variance for random models. Vol. II. Unbalanced data*, Birkhäuser Boston Inc., Boston, MA, 2005, Theory, methods, applications, and data analysis. MR 2152283 (2006i:62003)
- [Stu09] Bernd Sturmfels, *Open problems in algebraic statistics*, Emerging applications of algebraic geometry, IMA Vol. Math. Appl., vol. 149, Springer, New York, 2009, pp. 351–363. MR 2500471 (2010g:13047)